FACILE INTERROGATION OF HIGH-ORDER EPISTASIS BETWEEN DISTAL SITES USING NEXT-GENERATION SEQUENCING

Justin Yoo, Department of Chemical Engineering, University of California, USA
Justin_yoo@ucsb.edu
Patrick Daugherty, Serimmune Inc., 150 Castilian Drive, USA
Michelle O'Malley, Department of Chemical Engineering, University of California, USA

Deep mutational scanning (DMS) combines next-generation sequencing and protein engineering to construct sequence-function landscapes and rapidly identify fitness optima. Practical use of these landscapes requires identification of all mutations in each protein variant due to the potential effects of epistasis, the interdependence between residues resulting in non-additive phenotypes. This phenomenon plays an important role in protein evolution and is often a necessary step along the path towards protein fitness optima. However, current methods to assign distal mutations to their corresponding gene are work-intensive, costly, and introduce potential sources of error. To overcome these limitations, we introduce a method compatible with DMS that matches distal mutations to their corresponding gene without additional experimental steps. Using this approach to screen ~2,000,000 unique protein variants, we engineer a human G protein-coupled receptor with a 15-fold improvement in ligand binding affinity and observe prevalent epistasis between distal residues within the ligand binding pocket. Compared to variants containing only proximal substitutions, those harboring missense mutations in distal sites demonstrate significantly greater functional activity in our screen. This method can be applied immediately to all experiments using Illumina next-generation sequencing and provides a facile approach to illuminate complex mechanisms underlying key protein functions.
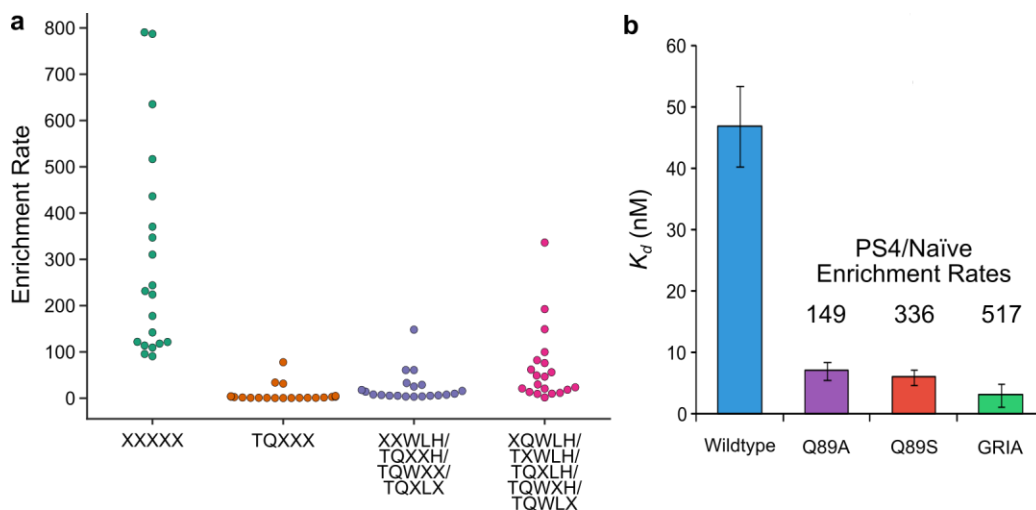
Figure 1 – (*a*) The top 20 GPCR variants containing at least two distal, mutated residues (XXXXX) exhibit far greater enrichment rates compared to the top 20 variants containing proximal triple (TQXXX), double (e.g. XXWLH), or single (e.g. XQWLH) substitutions. (*b*) Compared to wildtype A_2aR, highly enriched variants, Q89A/S and GRIA, bind $[^3H]$-NECA with ~7 – 15-fold greater affinity (K_d).